

In re Application of:
Frank Paetzold et al.
Application No.: 09/929,516
Filed: August 13, 2001
Page 2

PATENT
Docket No.: EYEM1340

Claim Listing:

1. (currently amended) Method for generating facial animation values using a sequence of facial image frames and synchronously captured audio data of a speaking actor, comprising the steps for:

providing a plurality of visual-facial-animation values based on tracking, ~~without using markers attached to the actor's face~~, of facial features in the sequence of facial image frames of the speaking actor;

providing a plurality of audio-facial-animation values based on visemes detected using the synchronously captured audio voice data of the speaking actor; and

combining the plurality of visual facial animation values and the plurality of audio facial animation values to generate output facial animation values for use in facial animation.

2. (original) Method for generating facial animation values as defined in claim 1, wherein the output facial animation values associated with a mouth for a facial animation are based only on the respective mouth-associated values of the plurality of audio facial animation values.

3. (currently amended) Method for generating facial animation values as defined in claim 1, wherein the output facial animation values associated with a mouth for a facial animation are based ~~only on~~ a weighted average of the respective mouth-associated values of the plurality of visual facial animation values and the respective mouth-associated values of the plurality of audio facial animation values.

4. (original) Method for generating facial animation values as defined in claim 1, wherein the output facial animation values associated with a mouth for a facial animation are based on Kalman filtering of the respective mouth-associated values of the plurality of visual facial animation values and the respective mouth-associated values of the plurality of audio facial animation values.

5. (original) Method for generating facial animation values as defined in claim 1, wherein the step of combining the plurality of visual facial animation values and the plurality of audio facial animation values to generate output facial animation values includes detecting whether speech is occurring in the synchronously captured audio voice data of the speaking actor and, while speech is detected as occurring, generating the output facial animation values associated with a mouth based only on the respective mouth-associated values of the plurality of audio facial animation values and, while speech is not detected as occurring, generating the output facial animation values associated with a mouth based only on the respective mouth-associated values of the plurality of visual facial animation values.

6. (original) Method for generating facial animation values as defined in claim 1, wherein the tracking of facial features in the sequence of facial image frames of the speaking actor is performed using bunch graph matching.

7. (original) Method for generating facial animation values as defined in claim 1, wherein the tracking of facial features in the sequence of facial image frames of the speaking actor is performed using transformed facial image frames generated based on wavelet transformations.

8. (original) Method for generating facial animation values as defined in claim 1, wherein the tracking of facial features in the sequence of facial image frames of the speaking actor is performed using transformed facial image frames generated based on Gabor wavelet transformations.

9. (currently amended) Apparatus for generating facial animation values using a sequence of facial image frames and synchronously captured audio data of a speaking actor, comprising:

means for providing a plurality of visual-facial-animation values based on tracking, ~~without using markers attached to the speaking actor's face,~~ of facial features in the sequence of facial image frames of the speaking actor;

means for providing a plurality of audio-facial-animation values based on visemes detected using the synchronously captured audio voice data of the speaking actor; and

means for providing a plurality of visual-facial-animation values based on tracking of facial features in the sequence of facial image frames of the speaking actor;

means for combining the plurality of visual facial animation values and the plurality of audio facial animation values to generate output facial animation values for use in facial animation.

10. (original) Apparatus for generating facial animation values as defined in claim 9, wherein the output facial animation values associated with a mouth for a facial animation are based only on the respective mouth-associated values of the plurality of audio facial animation values.

11. (original) Apparatus for generating facial animation values as defined in claim 9, wherein the output facial animation values associated with a mouth for a facial animation are based on a weighted average of the respective mouth-associated values of the plurality of visual facial animation values and the respective mouth-associated values of the plurality of audio facial animation values.

12. (original) Apparatus for generating facial animation values as defined in claim 9, wherein the output facial animation values associated with a mouth for a facial animation are based on Kalman filtering of the respective mouth-associated values of the plurality of visual facial animation values and the respective mouth-associated values of the plurality of audio facial animation values.

13. (original) Apparatus for generating facial animation values as defined in claim 9, wherein the means for combining the plurality of visual facial animation values and the plurality of audio facial animation values to generate output facial animation values includes means for detecting whether speech is occurring in the synchronously captured audio voice data of the speaking actor and, while speech is detected as occurring, generating the output facial animation values associated with a mouth based only on the respective mouth-associated values of the plurality of audio facial animation values and, while speech is not detected as occurring, generating the output facial animation values associated with a mouth based only on the respective mouth-associated values of the plurality of visual facial animation values.

14. (original) Apparatus for generating facial animation values as defined in claim 9, wherein the tracking of facial features in the sequence of facial image frames of the speaking actor is performed using bunch graph matching.

15. (original) Apparatus for generating facial animation values as defined in claim 9, wherein the tracking of facial features in the sequence of facial image frames of the speaking actor is performed using transformed facial image frames generated based on wavelet transformations.

16. (original) Apparatus for generating facial animation values as defined in claim 9, wherein the tracking of facial features in the sequence of facial image frames of the speaking actor is performed using transformed facial image frames generated based on Gabor wavelet transformations.

17. (new) Apparatus for generating facial animation values as defined in claim 9, wherein the tracking of facial features in the sequence of facial image frames of the speaking actor is performed without using markers attached to the speaking actor's face.

18. (new) Apparatus for generating facial animation values as defined in claim 11, wherein the output facial animation values are calculated using the following equation:

$$\left(\underline{f}_n = \frac{\underline{\sigma}_n^a}{\underline{\sigma}_n^a + \underline{\sigma}_n^v} \cdot \underline{a}_n + \frac{\underline{\sigma}_n^v}{\underline{\sigma}_n^v + \underline{\sigma}_n^a} \cdot \underline{v}_n \right)_i$$

where:

\underline{f}_n are the output facial animation values;

\underline{v}_n are the visual facial animation values;

\underline{a}_n are the respective mouth-associated values of the audio facial animation values;

σ_n^a are the weights for the audio facial animation values; and

σ_n^v are the weights for the visual facial animation values.

19. (new) Method for generating facial animation values as defined in claim 1, wherein the tracking of facial features in the sequence of facial image frames of the speaking actor is performed without using markers attached to the speaking actor's face.

20. (new) Method for generating facial animation values as defined in claim 3, wherein the output facial animation values are calculated using the following equation:

$$\left(\underline{f}_n = \frac{\underline{\sigma}_n^a}{\underline{\sigma}_n^a + \underline{\sigma}_n^v} \cdot \underline{a}_n + \frac{\underline{\sigma}_n^v}{\underline{\sigma}_n^v + \underline{\sigma}_n^a} \cdot \underline{v}_n \right)_i$$

where:

\underline{f}_n are the output facial animation values;

\underline{v}_n are the visual facial animation values;

\underline{a}_n are the respective mouth-associated values of the audio facial animation values;

$\underline{\sigma}_n^a$ are the weights for the audio facial animation values; and

$\underline{\sigma}_n^v$ are the weights for the visual facial animation values.